Wavelet-based Multiresolution Forecasting

Saif Ahmad, Ademola Popoola, and Khurshid Ahmad

Department of Computing, University of Surrey, Guildford GU2 7XH
Surrey, UK

Abstract – In this report, we discuss results of modelling and forecasting nonstationary financial time series using a combination of the maximal overlap discreet wavelet transform (MODWT) and fuzzy logic. A financial time series is decomposed into an over complete, shift invariant wavelet representation. A fuzzy-rule base is created for each individual wavelet sub-series to predict future values. To form the aggregate forecast, the individual wavelet sub-series forecasts are recombined utilizing the linear reconstruction property of the wavelet multiresolution analysis (MRA). Results are presented for IBM, NASDAQ and S&P 500 daily (adjusted) close values.

I. Introduction

The successful application of modern modelling techniques like neural networks and fuzzy logic to financial time series requires a certain uniformity (stationarity) of the data. Financial time series data is inherently nonstationary and may be a superposition of many sources exhibiting different dynamics. Neural networks (employing nonlinear autoregressions) and fuzzy logic models (employing fuzzy-rule bases) can be termed as global approximators where only one model is used to characterize an entire process. Therefore, such techniques usually give best results for stationary time series.

Recently, there has been an increased interest in multiresolution decomposition techniques like the wavelet transform for elucidating complex relationships in nonstationary financial time series . The wavelet transform can produce a good local representation of a signal in both time and frequency domain and is not restrained by the assumption of stationarity . Moreover, the wavelet approach has formalized old notions of decomposing a financial time series into trend, seasonal and business cycle components . Motivated by the spatial frequency resolution property of the wavelet transform, several hybrid schemes (local models) have been developed, for example , which combine wavelet analysis with machine learning approaches like neural networks for time series prediction.

In this report, we present results of predicting financial time series with a fuzzy-wavelet hybrid system that incorporates multiscale wavelet decompositions into a set of fuzzy-rule bases. The system employs a shift invariant wavelet transform called the maximal overlap discrete wavelet transform (MODWT) . Essentially, the so-called *à trous* filtering scheme is applied to generate MODWT decompositions of a financial time series. A fuzzy-rule base is created to predict each wavelet decomposition separately. To generate a global forecast, the prediction results of individual wavelet decompositions are combined directly using the linear reconstruction property of the wavelet multiresolution analysis (MRA).

Results are presented for three sets of data: (1) IBM daily prices from January 1982 to January 2004, (2) NASDAQ daily index values from October 1984 to June 2004, and (3) S&P 500 daily index values from January 1984 to June 2004.

II. Wavelet Based Pre-processing

A. Discrete Wavelet Transform (DWT)

The discrete wavelet transform (DWT) is a mathematical tool that projects a time series onto a collection of orthonormal basis functions (wavelets) to produce a set of wavelet coefficients. These coefficients capture information from the time series at different frequencies at distinct times. For a function f defined on the entire real line, a suitably chosen mother wavelet function ψ can be used to expand f as,

$$f(t) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} w_{jk} 2^{j/2} \psi\,(2^j t - k) \qquad (1)$$

where the functions $\psi\,(2^j t - k)$ are all orthogonal to one another. The coefficient w_{jk} conveys information about the behaviour of the function f concentrating on effects of scale around 2^{-j} near time $k \times 2^{-j}$.

The DWT can effectively compress a wide range of signals – a large proportion of DWT coefficients can actually be set to zero without appreciable loss of information. It can deal well with heterogeneous and transient behaviour that makes it so attractive for financial time series analysis. However, one problem associated with the application of the DWT for time series analysis is that it suffers from a lack of translation invariance. This means that circularly shifting a time series will not necessarily shift its DWT coefficients in a similar manner.

B. Maximal Overlap DWT (MODWT) and à trous Filtering

This problem can be tackled by means of a highly redundant non-orthogonal transform called the maximal overlap discrete wavelet transform (MODWT) . For a redundant transform like the MODWT, an N samples input time series will have an N samples resolution scale for each resolution level. Therefore, the features of wavelet coefficients in a multiresolution analysis (MRA) will be lined up with the original time series in a meaningful way.

For a time series X with arbitrary sample size N, the jth level MODWT wavelet ($\widetilde{W}_j$) and scaling ($\widetilde{V}_j$) coefficients are defined as,

$$\begin{aligned} \widetilde{W}_{j,t} &\equiv \sum_{l=0}^{L_j-1} \widetilde{h}_{j,l} X_{t-l \bmod N} \\ \widetilde{V}_{j,t} &\equiv \sum_{l=0}^{L_j-1} \widetilde{g}_{j,l} X_{t-l \bmod N} \end{aligned} \tag{2}$$

where $\widetilde{h}_{j,l} \equiv h_{j,l}/2^{j/2}$ are the MODWT wavelet filters, and $\widetilde{g}_{j,l} \equiv g_{j,l}/2^{j/2}$ are the MODWT scaling filters.

For a time series X with N samples, the MODWT yields an additive decomposition or MRA given by,

$$X = \sum_{j=1}^{J_0} \widetilde{D}_j + \widetilde{S}_{J_0} \tag{3}$$

where

$$\begin{aligned} \widetilde{D}_{j,t} &= \sum_{l=0}^{N-1} \widetilde{h}^{\circ}_{j,l} \widetilde{W}_{j,t+l \bmod N} \\ \widetilde{S}_{j,t} &= \sum_{l=0}^{N-1} \widetilde{g}^{\circ}_{j,l} \widetilde{V}_{j,t+l \bmod N} \end{aligned} \tag{4}$$

According to Equation (3), at a scale j, we obtain a set of coefficients $\{D_j\}$ each with the same number of samples (N) as in the original signal (X). These are called wavelet "details" and they capture local fluctuations over the whole period of a time series at each scale. The set of values S_{J0} provide a "smooth" or overall "trend" of the original signal. Adding D_j to S_{J0}, for $j = 1, 2, \ldots, J_0$, gives an increasingly more accurate approximation of the original signal. This additive form of reconstruction allows us to predict each wavelet sub-series (D_j, S_{J0}) separately and add the individual predictions to generate an aggregate forecast.

Since our intention is to make one-step-head predictions, we should perform the MODWT in such a way that the wavelet coefficients (for each level) at time point n should not be influenced by the behaviour of the time series beyond point n. That is we must perform the MODWT incrementally where a wavelet coefficient at a position n is calculated from the signal samples at positions less than or equal to n, but never larger. This will give us the flexibility of dividing the wavelet coefficients for training and testing and making one-step-ahead predictions in the same way as we would for the original signal. To accomplish this we make use of the time-based *à trous* filtering scheme proposed in

, which is briefly described as follows. Consider a signal $X(1), X(2), \ldots, X(n)$, where n is the present time point and perform the following steps:

1) For index k sufficiently large, carry out the MODWT transform (2), (3), and (4) on $\{X(1), X(2), \ldots, X(n)\}$.
2) Retain the coefficient values as well as the smooth values for the kth time point only: $D_1(k), D_2(k), \ldots, S_5(k)$. The summation of these values gives $X(k)$.
3) If $k < n$, set $k = k+1$ and go to Step 1.

This process produces an additive decomposition of the signal $X(k), X(k+1), \ldots, X(n)$, which is similar to the *à trous* wavelet transform decomposition on $X(1), X(2), \ldots, X(n)$. The above algorithm is further illustrated in Fig. 1 where we have shown a level-5 decomposition.

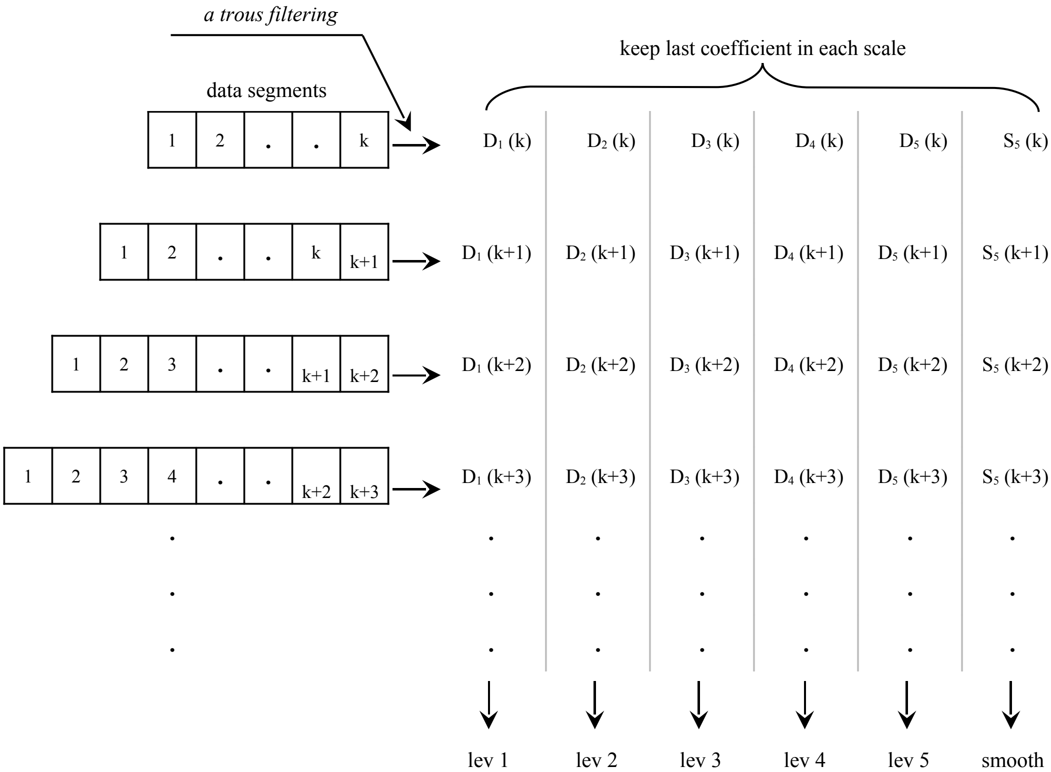

Fig. 1. Procedure for preparing data for *one-step-ahead* predictions using a fuzzy-wavelet hybrid model. Each time a segment of the time series is transformed using the MODWT, only the last coefficient is retained.

III. Hybrid Fuzzy-Wavelet Scheme for Time Series Prediction

Fig. 2 shows the proposed hybrid fuzzy-wavelet scheme for time series prediction. Given the time series $X(n)$, $n = 1, \ldots, N$, our aim is to predict the lth sample ahead, $X(N+l)$, of the series. That is $l = 1$ for single step prediction. This scheme basically involves three stages. In the first stage, the time series is decomposed into different scales using the MODWT. In the second stage, each scale is predicted by a separate fuzzy logic model and in the third stage the individual predictions at each scale are combined to generate an aggregate forecast.

The *time-based à trous* filtering scheme presented in Fig. 1, handles the temporal aspect of the data well to facilitate time series prediction. As an example, consider a financial index like the IBM prices with 109 data samples, on which we wish to carry out a level-3 *time-based à trous* transform. We can do this by implementing the scheme described in Fig. 1 by starting off with 10 samples ($k = 10$). That is, we simply carry out a level-3 MODWT on values $X(1)$ to $X(10)$. The last values of the wavelet coefficients at time-point $t = 10$ are kept because they are the most useful ones for prediction. Then we repeat the same procedure at time-point $t = 11$ (carry out a level-3 MODWT on values $X(1)$ to $X(11)$ and keep coefficients at time-point $t = 11$) and so on until we reach 109 (the total number of samples in the original data). In this manner, we will have wavelet decompositions for the time series from $t =$

10 to *t = 109*. The total number of samples in these decompositions will be 100, which is 9 less than the original time series since we chose *k = 10* to start our *à trous* filtering.

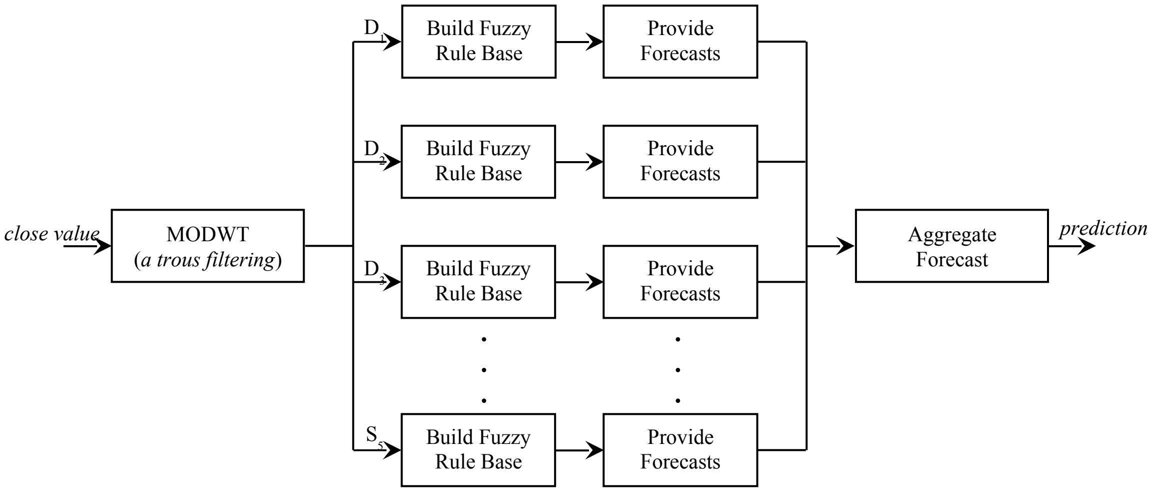

Fig. 2. Overview of the wavelet/fuzzy multiresolution forecasting system. D_1, ...D_5 are wavelet coefficients, S_5 is the signal "smooth" or "trend".

Fig. 3 shows the IBM prices from *t = 10* to *t = 109* (a total of 100 samples), and the corresponding (MODWT) wavelet transform computed by the above method. As the wavelet level increases, the corresponding coefficients become smoother. We will show in the next section that the ability of fuzzy models to capture dynamical behaviour varies with the wavelet resolution level.

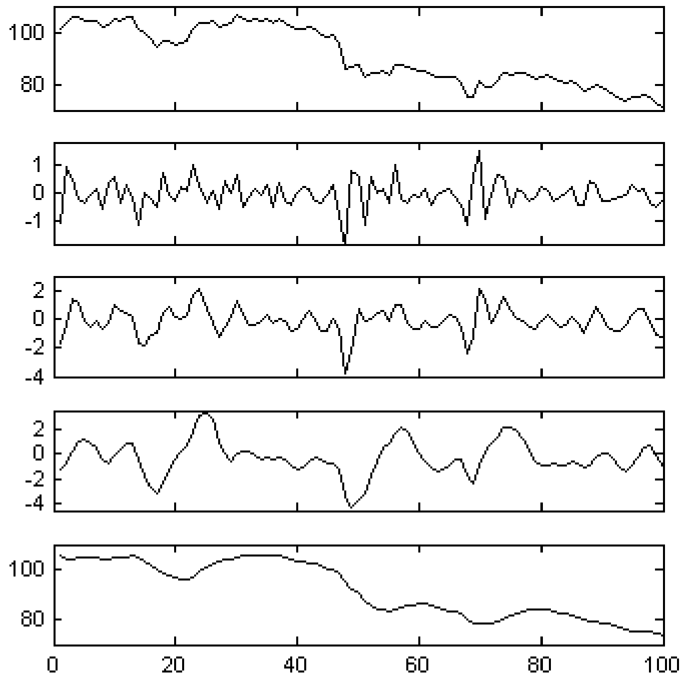

Fig. 3. Illustration of the *à trous* wavelet decomposition of IBM closing price series. From top to bottom: IBM closing price, D_1, D_2, D_3 and the wavelet "smooth".

In the second stage, the wavelet decompositions are partitioned into training and testing sets and separate fuzzy models are built for each, over the training set. Therefore, a *level-J* wavelet decomposition results in *J+1* fuzzy models:

$$
\begin{aligned}
&f_1 \quad : D_{1_{t-n}},\ D_{1_{t-n+1}}, \dots, D_{1_t} \ \rightarrow\ D_{1_{t+1}} \\
&f_2 \quad : D_{2_{t-n}}, D_{2_{t-n+1}}, \dots, D_{2_t} \ \rightarrow\ D_{2_{t+1}} \\
&. \\
&. \\
&f_J \quad : D_{J_{t-n}}, D_{J_{t-n+1}}, \dots, D_{J_t} \ \rightarrow\ D_{J_{t+1}} \\
&f_{J+1} \ : S_{J_{t-n}}, S_{J_{t-n+1}}, \dots, S_{J_t} \ \rightarrow\ S_{J_{t+1}}
\end{aligned}
\tag{5}
$$

We use the subtractive clustering method proposed by Chiu to build individual fuzzy-rule bases for the wavelet decompositions. A cluster radius of 0.5 is used for all predictions. A first-order (Takagi-Sugeno-Kang) TSK fuzzy inference system (FIS) is obtained on the training data, using a one-pass Matlab™ function without iterative optimisation. The FIS is then used on the test data for single step prediction. In the third stage, single step predictions for the wavelet decompositions are combined to give the next step forecast for time series *X*:

$$X_{t+1} = D_{1_{t+1}} + D_{2_{t+1}} + \dots + D_{J_{t+1}} + S_{J_{t+1}} \tag{6}$$

IV. Simulations and Performances

Analysis and simulations involved the daily (adjusted) closing value of three financial instruments: (1) IBM prices from January 1982 to January 2004, (2) NASDAQ index from October 1984 to June 2004, and (3) S&P 500 index from January 1984 to June 2004. To undertake a meaningful analysis, the IBM data was subdivided into 5 time series, each comprising roughly 5 years of data. Similarly NASDAQ and S&P 500 data was also subdivided into 4 time series, again each comprising roughly 5 years of data. This resulted in a total of 13 time series (each of length equal to around five years) to be analysed – 5 for IBM and 4 each for NASDAQ and S&P 500. For each of the 13 time series, a *training to testing ratio* of 80 to 20 per cent was employed (see Table I for more details). Fig. 4 shows one of the 13 time series analysed: the IBM daily close prices from January 2, 1987 to December 31, 1991.

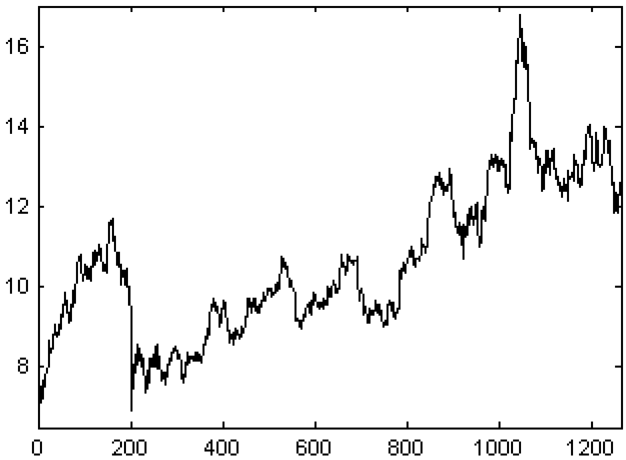

Fig. 4. Closing price for IBM from January 2, 1987 to December 31, 1991: a total of 1264 data samples.

We examined the approach of forecasting each wavelet derived coefficient series individually and then recombining the marginal forecasts. The objective was to perform one day ahead forecasting of the closing value for each time series and comparing the predictions with the target (original) series using the root mean squared error (RMSE) statistic, E_i (7).

$$E_i = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(P_{ij} - T_j)^2} \tag{7}$$

Each of the 13 time series was decomposed into five wavelet resolution levels using *à trous* filtering. The fuzzy system was trained separately on each wavelet sub-series (resolution) to generate one-step-ahead forecasts. Fig. 5 shows the one-step ahead predictions for each of the five wavelet coefficient series ($\{D_1\}$, $\{D_2\}$, $\{D_3\}$, $\{D_4\}$, $\{D_5\}$) and the "smooth" ($\{S_5\}$) of the IBM prices of Fig. 4 over a testing set comprising 252 days.

We observe that the ability of the fuzzy model to capture dynamical behaviour varies with the wavelet resolution level. The prediction is inaccurate at lower scales for which wavelet coefficients are noisy and irregular (top two series in Fig. 5). On the other hand, for higher scales, which are much more smooth and systematic, the prediction proceeds quite well (last four series in Fig. 5).

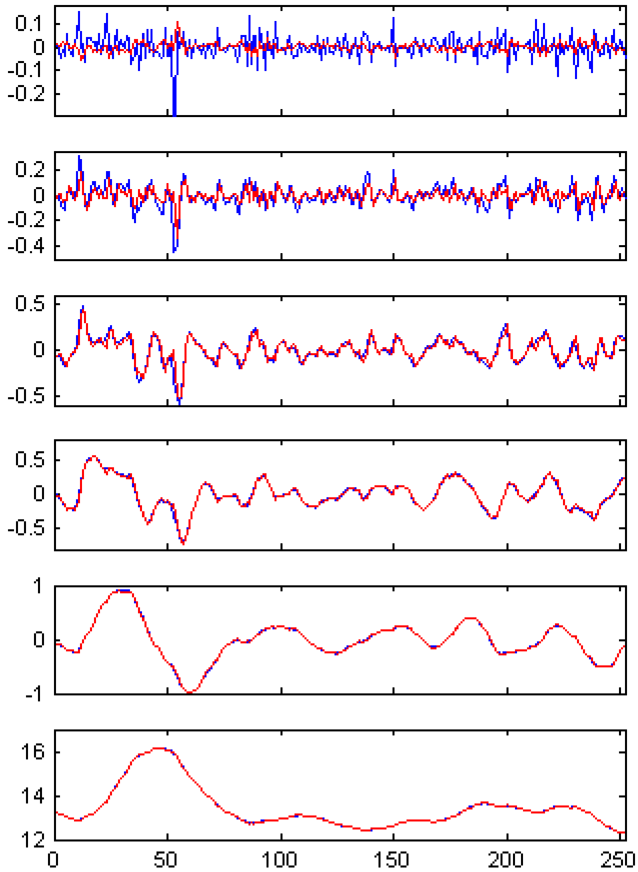

Fig. 5. From top to bottom: one step ahead predictions for the five wavelet coefficient series D_1, D_2, D_3, D_4, D_5 and the smooth series S_5, over a 252 days period on the testing set. The blue line is the target series while the red line is the prediction.

Fig. 6a shows the aggregate prediction (sum of the individual wavelet predictions of Fig. 5) for the IBM series over a test set of 252 samples. Fig. 6b shows the same prediction without the noisiest wavelet component (D_1: the first series in Fig. 5). It is clear that the fit to the data is much better in Fig. 6b where the noise component has been removed from the signal.

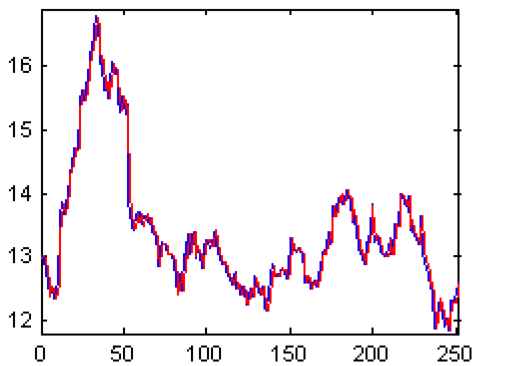

Fig. 6a. Aggregate prediction of IBM prices without de-noising over a 252 days period on the testing set. The blue line is the target series while the red line is the prediction.

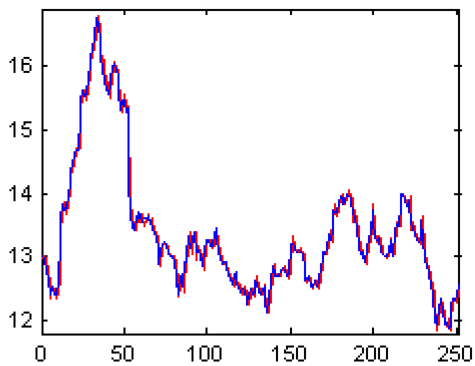

Fig. 6b. Aggregate prediction of IBM prices with de-noising over a 252 days period on the testing set. The blue line is the target series while the red line is the prediction.

We thus conclude that noise in the signal contributes significantly to the RMSE of prediction and systematic removal of noise improves the quality of the overall forecast. In TABLE I, we present a summary of the results for all the 13 time series that were analysed using the above methodology. The last two columns of TABLE I show the RMSE without and with wavelet de-noising respectively. We observe that the RMSE drops significantly for all the 13 time series when we exclude the noisy component (D_1) for training and testing the fuzzy prediction model.

TABLE I: RMSE for all the 13 time series with and without wavelet de-noising

Data	Time Series No.	Period		Samples	Train	Test	RMSE without De-noising	RMSE with De-noising
		From	To					
IBM 1	1	04-Jan-82	31-Dec-86	1264	1012	252	0.11	0.08
IBM 2	2	02-Jan-87	31-Dec-91	1264	1012	252	0.22	0.17
IBM 3	3	02-Jan-92	31-Dec-96	1265	1012	253	0.56	0.44
IBM 4	4	02-Jan-97	31-Dec-01	1257	1012	245	2.30	1.80
IBM 5	5	02-Jan-02	12-Jan-04	511	408	103	1.01	0.79
Nasdaq 1	6	11-Oct-84	30-Dec-88	1067	853	214	2.05	1.60
Nasdaq 2	7	03-Jan-89	31-Dec-93	1265	1012	253	4.97	3.88
Nasdaq 3	8	03-Jan-94	31-Dec-98	1263	1012	251	29.35	22.99
Nasdaq 4	9	04-Jan-99	30-Jun-04	1380	1104	276	22.87	17.90
S&P 1	10	03-Jan-84	30-Dec-88	1265	1012	253	2.85	2.22
S&P 2	11	03-Jan-89	31-Dec-93	1265	1012	253	2.43	1.90
S&P 3	12	03-Jan-94	31-Dec-98	1263	1012	251	13.66	10.71
S&P 4	13	04-Jan-99	30-Jun-04	1379	1103	276	8.33	6.51

V. AFTERWORD

In this report we have presented a method that combines shift invariant wavelet transform pre-processing with fuzzy logic for financial time series prediction. The results show significant advantages of wavelet pre-processing for time series analysis and prediction. Wavelets provide a formal method to de-noise, de-seasonalize, de-trend, and break down a complex time series into simpler units to facilitate accurate prediction. A significant improvement in the RMSE statistic is observed when the so-called *noise* component is removed from the signal (TABLE I). However, this is just one aspect or benefit of using wavelet pre-processing for time series prediction. Since the wavelet transform has the capability of decomposing a time series into the trend, seasonal and irregular components, appropriate prediction techniques can be applied to each component to gain an overall efficiency in forecast. Here we demonstrated the usefulness of a hybrid fuzzy-wavelet prediction scheme. But there are other methods to be explored, for example, fitting an autoregressive model to the "smooth" or "trend" component, a SARIMA model to the seasonal component, and a bootstrapping model to the *noise* component.

REFERENCES

[1] Aussem, A., J., and F. Murtagh. 1997. "Combining neural network forecasts on wavelet-transformed time series." *Connection Science* 9 (1): 113-121.

[2] Chiu, S. 1994. "Fuzzy Model Identification Based on Cluster Estimation." *Journal of Intelligent & Fuzzy Systems* 2 (3): 267-278.

[3] Chiu, S. 1997. "Extracting Fuzzy Rules from Data for Function Approximation and Pattern Classification." D. Dubois, H. Prade, and R. Yager (Eds.), Fuzzy Information Engineering: A Guided Tour of Applications. *John Wiley and Sons*.

[4] Gençay, R., F. Selcuk, and B. Whitcher. 2002. "An Introduction to Wavelets and Other Filtering Methods in Finance and Economics." *Academic Press*, New York.

[5] Mallat, S. 1989. "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11: 674-93.

[6] Percival, D. B., and A. T. Walden. 2000. "Wavelet Methods for Time Series Analysis." *Cambridge University Press*.

[7] Ramsey, J. B. 1999. "The contribution of wavelets to the analysis of economic and financial data." *Phil. Trans. R. Soc. Lond. A* 357: 2593-2606.

[8] Shensa, M. J. 1992. "The discrete wavelet transform: wedding the à trous and Mallat algorithms." *IEEE Trans. Signal Process.* 40: 2464-2482.